4K

УДК 004.81:004.932.001.8

А.А. Каргин [1]***, Е.Е. Пятикоп*** [2]
[1] Донецкий национальный государственный университет, г. Донецк, Украина
[2] Приазовский государственный технический университет, г. Мариуполь, Украина
Pjatikopelena@rambler.ru

Применение нечетких моделей когнитивных знаний восприятия изображения в задаче локализации строк текста

В статье описываются знания когнитивной психологии о первичной обработке информации в зрительной системе. Вводится математическое описание этих знаний с помощью нечетких моделей (модель ганглиозной клетки, модель клетки латерального коленчатого тела, модели клеток первичной зрительной коры). Приводятся результаты применения этих моделей для решения задачи локализации строк текста.

Введение

Проникновение компьютерного зрения в основные сферы человеческой деятельности привело к ряду задач, связанных с обработкой и анализом изображений, конечной целью которых является распознавание объекта. Решение последней задачи, в свою очередь, включает ряд таких этапов, как восприятие поля зрения, сегментация, нормализация выделенных объектов, распознавание [1]. При выполнении каждого из этапов применяются различные методы. В источниках [1-3] предложены классификации основных методов обработки и распознавания изображений. Среди них такие средства, как линейный дискриминантный анализ, факторный анализ, фильтры, метод главных компонент, скрытые Марковские модели, вейвлет-преобразования, нейронные сети и другие.

Каждый из методов имеет определенную область применения, которая зависит от характера различий входных и эталонных изображений, от помеховой обстановки в поле зрения, требований к объемам вычислений и скорости принятия решений. Объемы цифровой визуальной информации увеличиваются, меняются условия распознавания, и современные системы требуют более глубокого интеллектуального анализа. Поэтому для обработки изображения исследуется возможность применения методов когнитивных наук [4-6].

В когнитивной психологии на основе наблюдений за человеком за длительный период предложны и апробированы экспериментально модели организации практически всех когнитивных функций человека: восприятие зрительной, звуковой информации, организация памяти разных уровней [7-9].

Эти знания дают возможность моделировать основы биологического зрения и применять полученные модели для обработки изображения с учётом особенностей восприятия человека. В данной статье рассматривается использование когнитивных моделей для решения задачи локализации строк текста на изображении.

Постановка задачи

Задача локализации строк текста на странице изображения ставится следующим образом. Пусть имеется изображение текста в градациях серого цвета 8 бит. Необходимо определить ориентацию (угол наклона) и ширину строк текста, а также данные о местоположении строк на изображении. Для анализа изображения применить метод, основанный на знаниях когнитивной психологии.

Знания когнитивной психологии

Если бросить мимолетный взгляд на изображение, не фиксируя взгляд на деталях, то происходит восприятие наиболее простых (укрупненных) паттернов в рассматриваемой задаче строк текста, где строки – это чередующиеся полосы одного направления и приблизительно одной ширины. Так происходит, если изображение попадает в область периферического зрения, а также, если смотреть на изображение мимолетом или издалека. После проецирования изображения на сетчатку в ней происходят такие процессы:

1. Первые фоторецепторы – палочки и колбочки – через промежуточные биполярные клетки активизируют определенные рецептивные поля ганглиозных клеток, которые могут быть типа «ON» или «OFF». Для некоторых ганглиозных клеток, если свет попадает на маленькую область сетчатки, будет отмечаться увеличение частоты генерации разрядов (ON-центр). Но если свет падает на область вокруг этого чувствительного центра, самопроизвольная частота генерации разрядов снижается. Существуют также ганглиозные клетки, у которых самопроизвольная частота разрядов снижается, когда свет попадает в центр (OFF-центр), и увеличивается, когда свет падает в окружающую их область. Эти клетки называют еще релейными [7]. Множество активных ганглиозных клеток образует первую «проекцию» изображения.

2. Далее, большинство аксонов ганглиозных клеток образуют синаптические связи с клетками латерального коленчатого тела (ЛКТ), рецептивные поля которых очень похожи на рецептивные поля ганглиозных клеток сетчатки: имеют центральные «зоны включения» и периферийные «зоны выключения» или наоборот. Помимо этого клетки ЛКТ отличаются размерами и делятся на две группы: парвоцеллюлярные и магноцеллюлярные клетки [9].

3. В ЛКТ первая «проекция» изображения становится пульсирующей [10]. Сразу же после саккады[1] в ЛКТ формируется ретинотопическая карта, которая передается в первичную зрительную кору. Диаметр рецептивных полей уменьшается и на основе его формируются новые ретинотопические карты. До очередного скачка зрительная кора перерабатывает данные, полученные из ЛКТ.

4. Для возбуждения клеток зрительной коры требуется более тонкий механизм, что отражается на анатомических особенностях самих клеток и на сигналах, необходимых для их возбуждения. Существует несколько типов этих клеток: *простые* клетки реагируют только на линейные сегменты, ориентированные определенным образом; *сложные* клетки требуют движения в определенном направлении; *гиперсложные* клетки требуют, чтобы находящиеся внутри их рецептивных полей стимулы были определенной длины. Кроме этого известно [7], [9], что клетки, реагирующие на одну ориентацию, формируют собой колонку зрительной коры (рис. 1). При этом рецептивные поля

[1] Движения глаз, переводящие точку фиксации с одного участка изображения на другой, используемые преимущественно для обследования и изучения поля зрения, а также для того, чтобы образы селективно отобранных деталей визуальных стимулов оказались на центральной ямке.

прилегающих клеток имеют другие ориентационные предпочтения, которые изменяются постепенно. Таким образом, среди множества ориентационных колонок на полосы (строки) текста наиболее активно проявит себя одна колонка (или группа смежных) определенной ориентации.

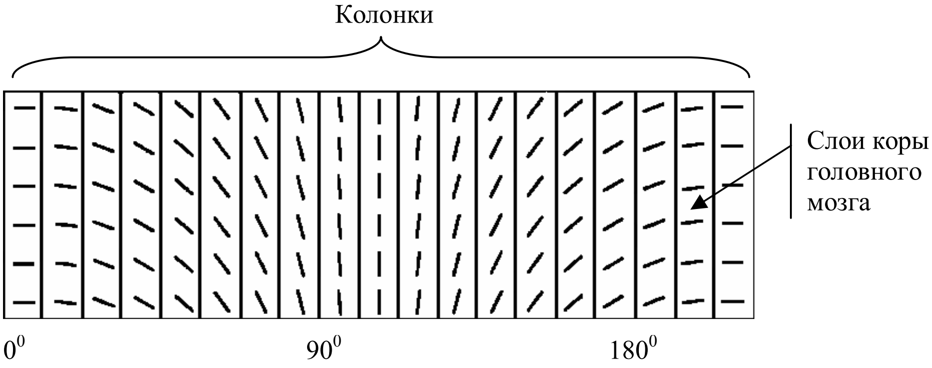


Рисунок 1 – Колонки клеток с разными ориентационными предпочтениями

Для применения этих знаний опишем их с помощью нечетких моделей.

Формальное описание моделей

Модель ганглиозной клетки

Представим исходное пиксельное поле *P* в виде множества клеток K_j. Каждая клетка сформирована на множестве пикселей $\{p_i\}$ начального изображения, которые характеризуются величиной яркости $f \in [0,1]$. Модель клетки, состоящая из ядра $Pk \in \{pk_i\}$ с радиусом *Rk* и полости $Pc \in \{pc_i\}$ с радиусом *Rc*, показана на рис. 2.

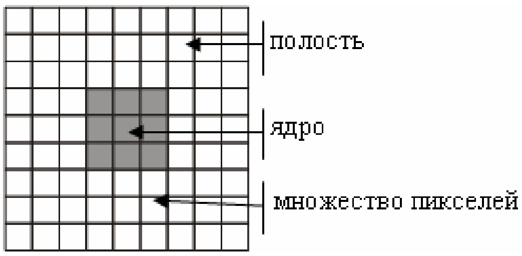


Рисунок 2 – Модель релейной клетки и ее характеристики

Для каждой из клеток K_j необходимо определить принадлежность к каждому из типов «ON» или «OFF». Поскольку фрагменты изображения не будут однозначно соответствовать приведенным типам, то принадлежность к конкретному типу будет иметь нечеткий характер. Тогда характеристиками каждой клетки являются функция уверенности $\mu_{\underset{\sim}{ON}}(\theta_i)$, которая показывает величину принадлежности K_j клетки к типу «включено» (ON), и функция уверенности $\mu_{\underset{\sim}{OFF}}(\varphi_i)$, которая показывает величину принадлежности K_j клетки к типу «выключено» (OFF). Функции принадлежности $\mu_{\underset{\sim}{ON}}(\theta_i)$ и $\mu_{\underset{\sim}{OFF}}(\varphi_i)$ определяются с помощью выражения (1):

$$\mu_{\underset{\sim}{ON}}(\theta_i) = \begin{cases} 0 & \text{для } \theta_i < 0{,}5; \\ \theta_i & \text{для } \theta_i \geq 0{,}5; \end{cases} \qquad \mu_{\underset{\sim}{OFF}}(\varphi_i) = \begin{cases} 0 & \text{для } \varphi_i < 0{,}5; \\ \varphi_i & \text{для } \varphi_i \geq 0{,}5, \end{cases} \tag{1}$$

где $\theta_i = \dfrac{(1-\chi_{Pk}) + \chi_{Pc}}{2}$, $\varphi_i = \dfrac{(1-\chi_{Pc}) + \chi_{Pk}}{2}$.

Величина χ_{Pk} показывает степень насыщенности ядра белым, а χ_{Pc} – степень насыщенности белым цветом полости и рассчитывается как

$$\chi_R = \frac{\sum_{p_i \in R} f(p_i) - n \cdot \min_{p_i \in R}\{f(p_i)\}}{\delta \cdot n},$$

где $R \in \{Pk, Pc\}$, n – количество пикселей области R, а величина δ отвечает за контрастность изображения и рассчитывается по формуле $\delta = \max_{p_i \in K_i}\{f(p_i)\} - \min_{p_i \in K_i}\{f(p_i)\}$.

Модель клетки ЛКТ

Математическая модель клеток данного уровня клеток – это есть выражение (1), но с изменением размеров Rk и Rc, следующим образом: $Rk' = \beta \cdot Rk$ и $Rc' = \beta \cdot Rc$, где для парвоцеллюлярных клеток $\beta < 1$, а для магноцеллюлярных клеток $\beta > 1$.

Модель клеток первичной зрительной коры

Поскольку для активизации клеток этого уровня необходима активность группы смежных релейных клеток [8], то введем понятие детектора. Совокупность смежных клеток показана на рис. 3.

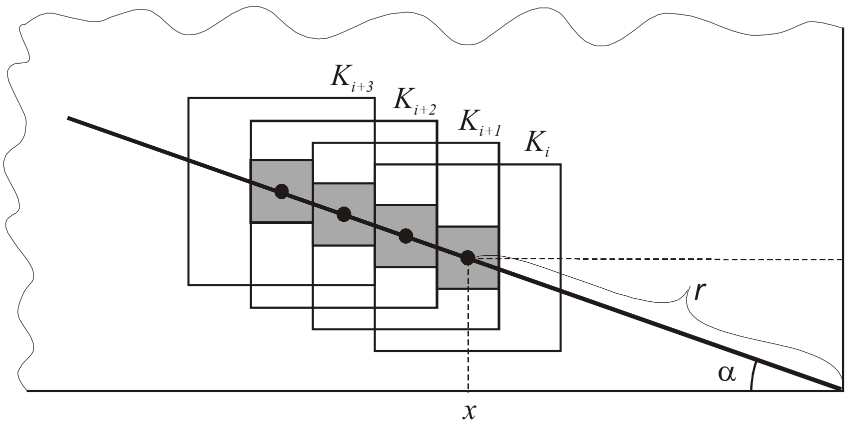


Рисунок 3 – Совокупность смежных клеток

Тогда модель детектора – это выражение

$$D^{\alpha} = < K^{\alpha,r}, \alpha, \omega, \ell^*, \delta(D^{\alpha}) >, \tag{2}$$

где $K^{\alpha,r} = < K^{\alpha,r+2Rk}, K^{\alpha,r+4Rk} \ldots K^{\alpha,r+\ell^* \cdot 2Rk} >$; α – угол ориентации от 0° до 175°; ω – ширина детектора, определяемая как ω = Rk; ℓ^* – длина детектора, определяемая количеством клеток, смежных $K^{a,r}$, но не менее 4; $\delta(D^{\alpha})$ – функция уверенности наличия детектора, вычисляемая по формуле:

$$\delta(D^{\alpha}) = \begin{cases} 0, & \text{если } \prod_{i \in 1 \ldots \ell^*} \mu(K_i) = 0 \\ \frac{1}{\ell^*} \cdot \sum_{i \in 1 \ldots \ell^*} \mu(K_i), & \text{если } \prod_{i \in 1 \ldots \ell^*} \mu(K_i) > 0, \end{cases} \tag{3}$$

где $\mu(K_i)$ вычисляется по формулам (1).

На одном проведенном луче может быть несколько детекторов. Пусть количество этих детекторов будет равно n_{α}. Тогда множество всех детекторов будет $D^{\alpha} = \{D_j^{\alpha}\},\ j \in 1 \ldots n_{\alpha}$, определяемых выражением (2).

На основе понятия детектора опишем модель различных видов клеток зрительной коры.

Так, простая клетка S^{α}, реагирующая только на определенную ориентацию, описывается выражением:

$$S^{a} =< K^{\alpha,r}, \alpha, \omega, \ell, \delta(S^{\alpha}) >, \qquad (4)$$

где $\ell \geq \ell^{*}$, а $\delta(S^{\alpha}) = \min\limits_{j \in 1 \ldots n_{\alpha}} \{\delta(D_{j}^{\alpha})\}$.

Сложные клетки затронуты не будут, поскольку они реагируют на движение, а в данной работе рассматривается обработка статического изображения.

Гиперсложные клетки зрительной коры так же, как и простые, чувствительны к направлению сигнала, но существенной их особенностью является реакция на определенную длину ℓ. Поэтому гиперсложная клетка $G^{\alpha,\ell}$ может быть описана выражением:

$$G^{\alpha,\ell} =< K^{\alpha,r}, \alpha, \omega, \ell, \delta(G^{\alpha,\ell}) >, \text{ где } \delta(G^{\alpha,\ell}) = \delta(D^{a}). \qquad (5)$$

В одном направлении α гиперсложных клеток $G^{\alpha,\ell}$ одной длины ℓ может быть количество m_{α}. Тогда множество всех гиперсложных клеток будет описано $G^{\alpha,\ell}$ = $= \{G_{j}^{\alpha,\ell}\},\ j \in 1 \ldots m_{a}$. Но также могут быть длины $\ell_{1}, \ell_{2}, \ldots \ell_{z}$. Тогда множество гиперсложных клеток, реагирующих на определенные длины, выражается так $G^{\alpha,\ell_{i}}$ = $= \{G_{j}^{\alpha,\ell_{i}}\},\ i \in 1 \ldots z,\ j \in 1 \ldots m_{a}$

В итоге можно построить модель изображения в виде множества ориентационных колонок, где каждая колонка описывается выражением:

$$C^{\alpha} = \left\{S^{\alpha}, \{G_{j}^{\alpha,\ell_{i}}\}_{\ell_{i}}\right\}, \alpha = 0,5,10...175. \qquad (6)$$

Модель представления изображения

Тогда модель «проекции» изображения, то есть ретинотопической карты описывается выражением:

$$M^{\omega} =< \{C^{\alpha}\}_{\omega}, \omega, \lambda(M^{\omega}) >, \qquad (7)$$

где α – угол ориентации от 0° до 175°; ω – ширина детектора, определяемая как: $\omega = Rk$; $\lambda(M^{\omega})$ – функция уверенности наличия наибольшего количества гиперсложных клеток наибольшей длины, вычисляемая по формуле:

$$\lambda(M^{\omega}) = [\frac{1}{2}(\lambda_{1} + \lambda_{2})]_{\alpha^{*}}, \qquad (8)$$

где λ_{1} – величина, отображающая, насколько все активные релейные клетки способствовали активизации гиперсложных клеток; λ_{2} – среднее значение уверенности всех гиперсложных клеток шириной ω; α^{*} – угол ориентации от 0° до 175°, выбранный на множестве ориентационных колонок C^{α}.

Тогда λ_{1} определяется как: $\lambda_{1} = \dfrac{n^{K^{\omega}}}{n^{G^{\alpha,\ell,\omega}}}$,

где $n^{K^{\omega}}$ – количество активных ганглиозных клеток K^{ω} с размером ядра $Rk = \omega$; $n^{G^{\alpha,\ell,\omega}}$ – количество активных ганглиозных клеток K^{ω} с размером ядра $Rk = \omega$, активизировавших гиперсложные клетки $\{G^{\alpha,\ell_{i}}\}_{\ell_{i}}$.

В свою очередь, λ_{2} определяется как $\lambda_{2} = \dfrac{1}{m_{\alpha}} \sum\limits_{j} \sum\limits_{i} \delta(G_{i}^{\alpha,\ell_{j},\omega})$,

где m_{α} – общее количество гиперсложных клеток одного направления, а угол α^{*} определяется из выражения $l_{\alpha^{*}} = \max\limits_{\omega}\{\max\limits_{\alpha}\{l_{\alpha}\}\}$,

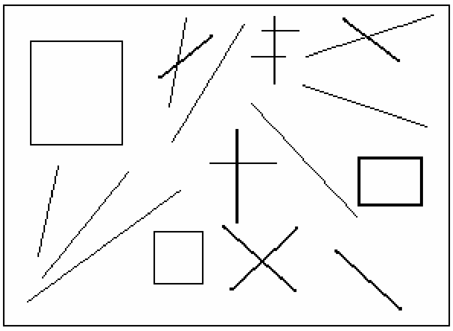
где $l_{\alpha} = \dfrac{1}{m_{\alpha}} \sum\limits_{j} \sum\limits_{i} G_{i}^{\alpha,\ell_{j},\omega}$ определяется для α от 0° до 175°.

После определения для каждой ретинотопической карты M^{ω_i} ее функции уверенности нужно выбрать карту с наибольшей уверенностью. Тогда гиперсложные клетки $G^{\alpha,\ell,\omega}$ этой карты будут соответствовать полосам шириной ω под углом α на изображении, то есть предполагаемым строкам.

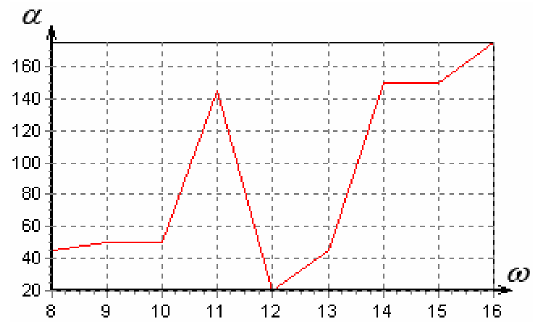
Результаты исследований и выводы

Для проведения экспериментов локализации строк текста было подобрано около 80 графических файлов формата BMP с градацией серого 8 бит. Среди тестируемых файлов были фрагменты текста отсканированных газет, журналов, книг, а также спам-изображения и файлы, полученные как копии экрана. Высота строк на изображениях колебалась от 8 до 35 пикселей, угол наклона – от 0° до 175°.

На рис. 4а показан пример изображения, на котором нет преобладания линий одной ориентации, поэтому на рис. 4б график демонстрирует скачкообразную динамику и дальнейший поиск нецелесообразен.



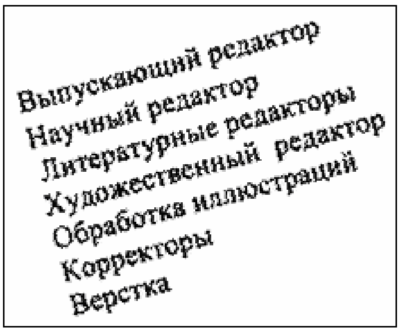
а) исходное изображение № 2



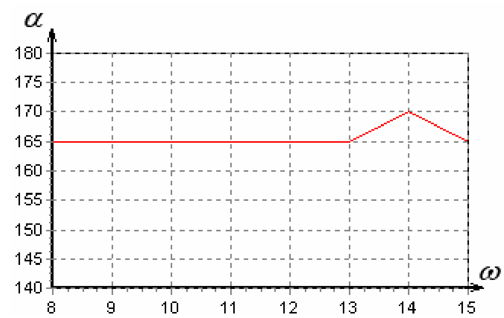
б) неоднозначный выбор приоритетной ориентации

Рисунок 4 – Примеры применения предлагаемых моделей

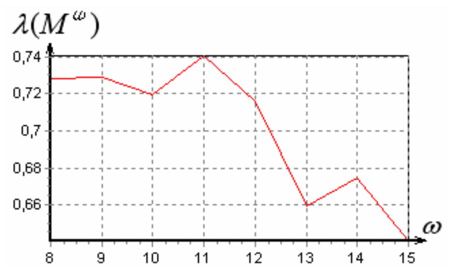
На рис. 5а представлен пример изображения со строками под углом 165°. График на рис. 5б показывает, что практически при всех размерах клетки преобладает приоритет ориентации 165°. Дальнейшее моделирование направлено на определение оптимальной ширины полос (высоты строк). На рис. 5в видно, что наибольшая уверенность полос получается при ширине 11 пикселей. На рис. 5г показано покрытие изображения клетками согласно выбранной ориентации и ширины. На основе данных о местоположении клеток следует локализация предполагаемых строк. Другие эксперименты по реализации данной модели приведены в [11].



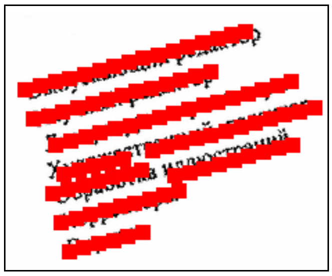
а) исходное изображение № 1



б) однозначный выбор приоритетной ориентации



в) однозначный выбор оптимальной ширины



г) результат локализации

Рисунок 5 – Примеры применения предлагаемых моделей

Таким образом, по результатам работы можно сделать следующие выводы:

– формализованы этапы первичной обработки информации в зрительной системе с помощью нечетких моделей;
– описана модель представления изображения текста со строками;
– введены и апробированы характеристики определения ориентации и ширины строк.

Литература

1. Путятин Е.П. Обработка изображений в робототехнике / Е.П. Путятин, С.И. Аверин. – М. : Машиностроение, 1990. – 320 с.
2. Гарячевская И.В. Автоматизация процесса разработки и отладки алгоритмов обработки изображений для СТЗ / И.В. Гарячевская, А.Я. Куземин // Искусственный интеллект. – 2004. – № 2. – С. 269-273.
3. Дудкин А.А. Обработка изображений и идентификация объектов в системах технического зрения / Р.Х. Садыхов, А.А. Дудкин // Искусственный интеллект. – 2006. – № 3. – С. 634-643.
4. Ярмошевич Е.И. Функциональная спектральная пространственно-временная модель формирования изображений объектов зрительной системой человека [Электронный ресурс] / Е.И. Ярмошевич, Е.Е. Михайлова, М.А. Пономаренко // Вестник ВГУ. Серия Системный анализ и информационные технологии. – 2008. – № 1. – С. 74-78. – Режим доступа к статье : www.vestnik.vsu.ru/pdf/analiz/2008/01/yarmoshevich.pdf
5. Bednar J.A. Scaling Self-Organizing Maps To Model Large Cortical Networks [Электронный ресурс] / James A. Bednar, Amol Kelkar and Risto Miikkulainen // Neuroinformatics. – 2004.– № 2. – P. 275-302. – Режим доступа : http://nn.cs.utexas.edu/keyword?bednar:neuroinformatics04 – название экрана.
6. Гладилин С.А. Нейронная сеть, воспроизводящая выходной сигнал ганглиозной клетки [Электронный ресурс] / С.А. Гладилин, Д.Г. Лебедев // Информационные процессы. – 2005. – Т. 5, № 3. – С. 258-264. – Режим доступа к статье : http://www.jip.ru/2005/258-264.pdf.
7. Андерсон Дж. Когнитивная психология / Дж. Андерсон. – [5-е изд.]. – СПб. : Питер, 2002. – 496 с. – (Серия «Мастера психологии»).
8. Солсо Р.Л. Когнитивная психология / Р.Л. Солсо. – СПб. : Питер, 2002. – 592 с. – (Серия «Мастера психологии»).
9. Шиффман Х.Р. Ощущение и восприятие / Х.Р. Шиффман. – [5-е изд.]. – СПб. : Питер, 2003. – 928 с. – (Серия «Мастера психологии»).
10. Демидов В.Е. Как мы видим то, что видим / В.Е. Демидов. – [2-е изд.]. – М. : Знание, 1987. – 240 с. – (Наука и прогресс).
11. Пятикоп Е.Е. Некоторые результаты компьютерных экспериментов локализация строк текста на основе когнитивных моделей восприятия изображения / Пятикоп Е.Е. // Вісник Донецького університету, Серія А : Природничі науки. – 2008. – № 2. – Ч. 2. – С. 527-532.

А.О. Каргін, О.Є. П'ятикоп

Застосування нечітких моделей когнітивних знань сприйняття зображення в задачі локалізації рядків

У статті описуються знання когнітивної психології про первинну обробку інформації в зоровій системі. Вводиться математичний опис цих знань за допомогою нечітких моделей (модель гангліозної клітки, модель клітки латерального колінчастого тіла, моделі кліток первинної зорової кори). Наводяться результати застосування цих моделей для вирішення задачі локалізації рядків тексту.

A.A. Kargin, E.E. Pyatikop

Application of Fuzzy Models of Cognitive Knowledge Perception of an Image in Task of String Localization

The article describes the knowledge of cognitive psychology on the primary information processing in the visual system. A mathematical description of this knowledge by using fuzzy models (model ganglion cells, cells of the model lateral geniculate nucleus, cells of a model primary visual cortex is entered). The results of applying these models to solve problem of line localization in text are presented.

Статья поступила в редакцию 13.03.2009.